A contribution to the study of the economic causes and consequences of climate change:

Estrada Porrua, F.

2015

14 Revisiting a flawed downscaling methodology

The methods employed by Magaña et al. (2012) in their analysis of the effects of climate change in northern Mexico are flawed. Our present analysis confirms the results in Estrada et al. (2012a), demonstrating that the proposed downscaling method is unacceptable for producing regional climate change scenarios. The method produces random spatial patterns and magnitudes that replace climate signals produced by general circulation models as well as those contained in the observed local-scale data. Consequently, any application that uses these regional scenarios should be revised and not used for supporting decision-making. The statistical downscaling literature provides a range of alternative methods that, if applied correctly, can be appropriate for producing regional climate change scenarios.

14.1 Introduction

The application of Model Output Statistics (MOS) originally proposed in non peer reviewed publications by Magaña & Caetano (2007), Zermeño (2008), INE-SEMARNAT (2009) and Magaña (2010) (hereafter collectively MCZ) was shown to be fundamentally flawed by Estrada et al. (2012a). Specifically, MCZ used the Climate Predictability Tool (CPT; http://iri.columbia.edu) with the 20th Century Climate Experiment (20c3m) simulations as predictors, and observed climate variables as predictands, for constructing regional climate change scenarios for Mexico.

Magaña et al. (2012; hereafter MZN[66]) again erroneously used the same method as MCZ, but with one main difference: an increase from a single to multiple predictors for specifying the transfer function. By extending the analysis in Estrada et al. (2012a), we show here that the modified method in MZN consists in multiplying each of the GCM simulations by a zero-mean random field, and consequently the resulting individual

---

This chapter was published as Estrada F, Livezey RE, Martínez-López B, Gay-García C (2013) Revisiting a flawed downscaling methodology: Comment on Magaña et al. (2012). Clim Res 56, 81-90. DOI 10.3354/cr01154

[66] MZN did not reference Estrada et al. (2012), despite its obvious relevance and online accessibility. Estrada et al. (2012) was published online in June 2011, while MZN was not accepted for publication in final form until January 2012.

scenarios as well as the transformed ensemble of models runs and its statistics are physically meaningless. The MZN method, like the MCZ method, is unacceptable for downscaling purposes, as it is based on a poor understanding of basic concepts of both climate variability and change and statistics.

The downscaling method in MCZ was used for Mexico's National Climate Change documents, to support the Mexican government's climate policy, e.g. the Fourth National Communication to the UN Framework Convention on Climate Change (INE-SEMARNAT 2009), the Fourth National Report to the Convention on Biological Diversity (CONABIO & SEMARNAT 2009) and a government report on the economics of climate change in Mexico (SEMARNAT & SHCP 2009). The resulting regional climate change scenarios are available at the Instituto Nacional de Ecología (National Ecology Institute of Mexico: http://zimbra.ine.gob.mx/escenarios/). The application of the erroneous method of MCZ has had a large impact in Mexico, because the credibility of the aforementioned national documents has been compromised by its use (Rosen 2010, Estrada et al. 2012a).

There is a variety of statistical methods in the downscaling literature that can be most properly applied for generating regional climate change scenarios that avoid fundamental errors such as those in MZC/MZN. The interested reader is referred to Wilby et al. 2004, Benestad et al. 2008, Maraun et al. 2010, Vrac et al. 2007, among others.

14.2 Methodology

14.2.1 Data and methods

The MCZ/MZN methodology is based on the MOS downscaling approach (Glahn & Lowry 1972) and was implemented by means of the CPT, an automated statistical downscaling toolbox designed for seasonal prediction based on canonical correlations and principal component (PC) linear regression. The MOS downscaling approach consists in estimating a statistical relationship between an observed local predictand and one or more large scale predictors that are the output of a dynamical model at some projection time. This relationship is applied to model output to estimate the projected values at local scales.

The downscaling method in MZN is applied for generating regional scenarios for Mexico, where the predictands are the Climate Research Unit (CRU) TS3.0 0.5º×0.5º gridded database of observed temperature and precipitation fields (Mitchell et al. 2004, Mitchell & Jones 2005) and the predictors are the first 5 PCs obtained from a variety of General Circulation Model (GCM) runs from the 20c3m experiment produced for the IPCC's Fourth Assessment Report (Christensen et al. 2007a). The PCs were estimated for a region encompassing Mexico and a southern portion of the USA. The statistical model chosen for estimating the downscaling relationships was PC linear regression.

Here we test the basic assumptions that would need to hold in order to make the method consistent and useful. For this purpose, we selected the precipitation fields from the same observational database used in MZN as the predictand variables, and the predictor variables are the first 5 PCs estimated from each of the 5 GISS-EH 20c3m precipitation fields (model runs are available at: http://www-pcmdi.llnl.gov/ipcc/about_ipcc.php). As in the MZN, the statistical model used for downscaling was multiple linear regression and the calibration period was 1901–1970. Examples shown here are for July, but very similar results are found for the other months of the year, as well as for other models and variables, including temperature (see Estrada et al. 2012a).

14.2.2 Principal components

A necessary condition for the proposed method to be consistent is that different GCM runs under the same 20c3m forcing scenario share a similar climate change signal detectable over the noise produced by the climate model. The MZN method uses PCs for separating the different modes of variability and for extracting the signal that could be associated with climate change. If this common climate change signal, imparted by the 20c3m radiative forcing, is indeed present and is strong enough to be detectable in at least 1 of the 5 sets of PCs, then the members of the sets representing the common signal should be well correlated. In this case, no matter which GCM simulation is used for downscaling, a potentially significant relation of roughly similar magnitude and sign may be obtained.

However, if the internal variability produced by the GCM dominates the sets of PCs, then their members will be independent from run to run. In this case, the method cannot be consistent, leading to patterns and magnitudes that are dependent on the model run that is used. Furthermore, since the PCs would mainly represent the climate model's noise — which is (1) dependent on the initial conditions and can be regarded as practically random, and (2) not correlated with the observed 20th century climate variability, due to the 20c3m experimental design — then the estimated statistical relationships will be meaningless for downscaling purposes. The resulting spatial patterns and magnitudes will be determined by the GCM's random internal noise, not by a systematic signal representing climate change. Two direct consequences would be (1) the estimated coefficients of the regression models will not be statistically different from zero, and thus the sign and magnitude shown by a particular point estimate of the coefficient are meaningless; (2) the differences in magnitudes and spatial patterns obtained from one model run to the other will be also random.

We computed all correlations between PCs for one run with another. Those between like-numbered PCs and their p-values are tabulated in Table 14.1. The series within each of the 5 sets of like-numbered PCs of the GISS-EH 20c3m runs can be considered linearly independent. Only 3 of the 50 correlation coefficients are statistically significant, as expected by chance at the 5% level; 2 of them correspond to the PC4 set. Of course, a climate change signal can be represented by a different numbered PC in different runs. Of the 300 correlations between all 5 PCs from the 5 different model runs, only 13 were statistically significant at the 5% level (Table 14.2). The number of statistically significant correlation coefficients does not increase even when rotated PCs are considered (varimax rotation), a variation that would allow for a clearer extraction of the climate change signal.

Table 14.1. Correlation matrices for the sets of the first 5 principal components (PCs) of the GISS_EH 20c3m model runs for July.

	Run 1	p	Run 2	p	Run 3	p	Run 4	p	Run 5
PC1									
Run 1	1								
Run 2	0.173	0.153	1						
Run 3	-0.094	0.439	0.037	0.759	1				
Run 4	-0.131	0.279	0.050	0.684	-0.033	0.786	1		
Run 5	-0.050	0.682	-0.042	0.732	0.066	0.587	-0.194	0.107	1
PC 2									
Run 1	1								
Run 2	-0.148	0.221	1						
Run 3	-0.117	0.336	-0.164	0.176	1				
Run 4	-0.132	0.275	-0.063	0.606	0.156	0.196	1		
Run 5	0.079	0.517	0.129	0.287	-0.168	0.164	-0.054	0.659	1
PC 3									
Run 1	1								
Run 2	-0.133	0.272	1						
Run 3	-0.060	0.621	*-0.211*	0.078	1				
Run 4	0.049	0.688	-0.088	0.470	-0.072	0.555	1		
Run 5	**-0.248**	0.039	0.154	0.201	0.000	0.999	0.084	0.490	1
PC 4									
Run 1	1								
Run 2	0.114	0.347	1						
Run 3	**0.280**	0.019	0.111	0.359	1				
Run 4	0.123	0.309	0.039	0.747	0.165	0.170	1		
Run 5	**-0.413**	0.000	-0.137	0.260	*-0.229*	0.056	0.068	0.576	1
PC 5									
Run 1	1								
Run 2	0.005	0.965	1						
Run 3	-0.093	0.445	-0.021	0.863	1				
Run 4	-0.121	0.320	0.044	0.717	-0.188	0.120	1		
Run 5	-0.115	0.345	0.074	0.545	0.112	0.358	0.177	0.142	1

Bold: $p \leq 0.05$; italic: $p \leq 0.1$

If these time series are independent realizations of the same data generating process (Estrada et al. 2012a), which, if any, should be used for statistical downscaling? The main building block of this method is that each of the temperature and precipitation simulations contains a clear enough climate change signal, imparted by the exogenous radiative forcing, and that this signal is detected above the GCM noise; another implied assumption is that all GCMs correctly represent the observed climate change signal even for small domains such as the one chosen in MZN. If the PCs obtained from the GCM are linearly independent across runs, this condition cannot be satisfied; the results of the downscaling method will depend on the realization that is used, and the value of the estimated coefficients is meaningless for downscaling purposes.

Table 14.2. Significant correlation coefficients from all possible combinations of the first 5 principal components (PCs) from the 5 runs of the GISS_EH 20c3m. Significant correlations: 4.3% (13/300) at $p \leq 0.05$

Combinations	Correlation coefficient	p
PC3_Run3, PC1_Run2	-0.257	0.032
PC4_Run4, PC1_Run5	0.260	0.030
PC4_Run5, PC2_Run1	-0.256	0.032
PC3_Run3, PC2_Run2	0.366	0.002
PC3_Run1, PC2_Run2	-0.276	0.021
PC3_Run5, PC3_Run1	-0.248	0.039
PC4_Run4, PC3_Run1	0.243	0.042
PC5_Run1, PC3_Run3	-0.305	0.010
PC5_Run5, PC3_Run3	0.264	0.027
PC4_Run3, PC4_Run1	0.280	0.019
PC4_Run5, PC4_Run1	-0.412	<0.001
PC5_Run2, PC4_Run4	0.298	0.012
PC5_Run4, PC4_Run4	0.459	<0.001

To further test for the presence of a common climate change signal in the PCs, we considered the following regression model:

$$y_{i,t}^{r1} = \alpha + \beta_1 y_{i,t}^{r2} + \beta_2 y_{i,t}^{r3} + \beta_3 y_{i,t}^{r4} + \beta_4 y_{i,t}^{r5} + \varepsilon_t, \qquad t = 1901,...,1970 \qquad (14.1)$$

where $y_{i,t}$ is the i-th PC (*i=1,...,5*) obtained from the GISS-EH 20c3m Runs 1 to 5; ε_t is a sequence of independent and identically distributed random errors such that $\varepsilon_t \sim N(0,\sigma^2)$; and α, β_1 to β_4 are unknown parameters to be estimated for each of the five Eqs. (14.1). If like-numbered PCs share common systematic information that could be associated with the climate change signal, and if this signal is detectable above the model's internal noise, then the 4 realizations of the i-th PC in Eq. (14.1) that are used as predictors would be expected to significantly contribute to explain the other realization of the same i-th PC $y_{i,t}^{r1}$ chosen as the dependent variable.

As expected from the results in Table 14.1, none of the slope coefficients is statistically significant, with the exception of the β_4 coefficient (but not β_2) that corresponds to the set of PC4s, and of the β_4 coefficient that corresponds to the set of PC3s (at the 10% level; Table 14.3). Since the coefficient estimates are not statistically different from zero, their signs and magnitudes have no meaning and depend on the particular sample. As before, these results suggest that the PCs are dominated by the model's internal variability, and no meaningful relationship between them and observed series can be expected for downscaling purposes.[67]

Table 14.3. Estimations of the slope coefficients in regression (14.1), using each of the first 5 principal components (PCs) estimated from the GISS-EH model 20c3m Run 1 as the predictand variable and the corresponding PC from the other 4 model simulations as the predictor variables.

	β_1	p	β_2	p	β_3	p	β_4	p
$y_{1,t}^{r1}$	0.129	0.136	-0.082	0.401	-0.113	0.206	-0.047	0.593
$y_{2,t}^{r1}$	-0.179	0.136	-0.124	0.359	-0.134	0.320	0.069	0.534
$y_{3,t}^{r1}$	-0.093	0.384	-0.070	0.519	0.051	0.660	*-0.206*	0.057
$y_{4,t}^{r1}$	0.037	0.725	0.182	0.145	0.135	0.289	**-0.352**	0.002
$y_{5,t}^{r1}$	0.016	0.905	-0.111	0.396	-0.133	0.521	-0.021	0.863

Bold: $p \leq 0.05$; italic: $p \leq 0.10$

These results agree with the downscaling literature and with expectations when using climate simulations such as 20c3m. Application of the MOS approach for developing bias-corrected high-resolution climate change scenarios is limited by the availability of long series of hindcasts (not unconstrained climate simulations) for calibrating the statistical model and establishing the relationships between model and observed variables; until now, MOS has been used only with 'nudged' climate simulations (which use data assimilation) or reanalysis data (Maraun 2010, Eden 2012).

As stated by Maraun et al. (2010, p. 13), “the GCM simulations for the 20th and 21th century do not represent the real temporal evolution of large−scale weather states in the past. ... For this reason, MOS has been applied so far to nonreanalysis GCMs only in the context of seasonal prediction, where the simulated and true atmospheric circulation partly match”. Unconstrained climate simulations of historical periods (such as 20c3m) are not intended to reproduce the observed evolution of internal atmospheric variability, and the GCM's random, internally generated variability component will be independent from the observed circulation and climate variable fields. Furthermore, this random component dominates daily to interannual time scales and is still substantial at decadal time scales (Eden et al. 2012). Consequently, comparisons of time series of daily,

[67] Tripling the potential predictors in (14.1) by including adjacent PCs or screening for predictors only sets the bar higher for significant predictors. The results in Tables 14.1 and 14.2 do not support pursuing these alternatives.

monthly, or annual precipitation between observations and freely evolving climate simulations are not relevant (Eden et al. 2012), and meaningful and consistent statistical relationships between freely-evolving simulations and observed climate variables cannot be derived. Under these circumstances, it makes no sense to use anything other than the climate change signal as predictor, and only if its signal-to-noise ratio is sufficiently large. In MZN this is further complicated because the climate change signal during the calibration period (1901–1970) and over the domain may be particularly weak in comparison to climate variability, and may not even have the same signature as the post-1975 signal.

14.2.3 Results of the MZN method

14.2.3.1 Results of a 'perfect climate model'

The least that should be asked for a statistical downscaling methodology is that the spatial patterns and magnitudes it produces are not a statistical artifact. If the method cannot fulfill this requirement in a reliable manner, its results would amount to trading a physically consistent and based climate change scenario for meaningless patterns and magnitudes produced by statistically inadequate models.

Following Estrada et al. (2012a), we tested if the MZN method produces consistent spatial patterns and magnitudes by investigating the behavior of the slope coefficients in the downscaling transfer function under the extreme assumption of having a perfect climate model.

Let us assume that GISS-EH 20c3m Run 1 is the 'observed' precipitation field over the region during the 20th century and choose the first 5 PCs obtained from one of the other 4 GISS-EH 20c3m runs as the predictor variables. Thus, 'observed' and simulated variables are realizations of the same data generating process, sharing the same scale, variable, model physics and external radiative forcings. This downscaling exercise is equivalent to assuming that the climate model is indeed perfect.

We considered the following regression equation:

$$P_{i,j,t}^{obs} = \alpha + \beta_1 P_{r,t}^{pc1} + \beta_2 P_{r,t}^{pc2} + \beta_3 P_{r,t}^{pc3} + \beta_4 P_{r,t}^{pc4} + \beta_5 P_{r,t}^{pc5} + \varepsilon_t \qquad (14.2)$$

where $P_{i,j,t}^{obs}$ is the precipitation field from Run 1 of the GISS-EH 20c3m for the coordinate i, j; $P_{r,t}^{pc1}, P_{r,t}^{pc2}, P_{r,t}^{pc3}, P_{r,t}^{pc4}, P_{r,t}^{pc5}$ represent, respectively, the first 5 PCs of the simulated precipitation fields obtained from the $r = 2,\ldots,5$ GISS-EH 20c3m runs; ε_t is a sequence of independent and identically distributed random errors such that $\varepsilon_t \sim N(0, \sigma^2)$; and α, β_1 to β_5 are unknown parameters to be estimated.

The results produced by the MZN method (figures 14.1 and 14.2) for β_1 show that the spatial patterns and magnitudes vary among runs, so that very different conclusions regarding future climate, impacts, vulnerability and adaptation studies could be reached in each case. The differences do not have a physical interpretation related to any climate change signal, but to the model's internal variability and to the initial conditions chosen for a particular run.

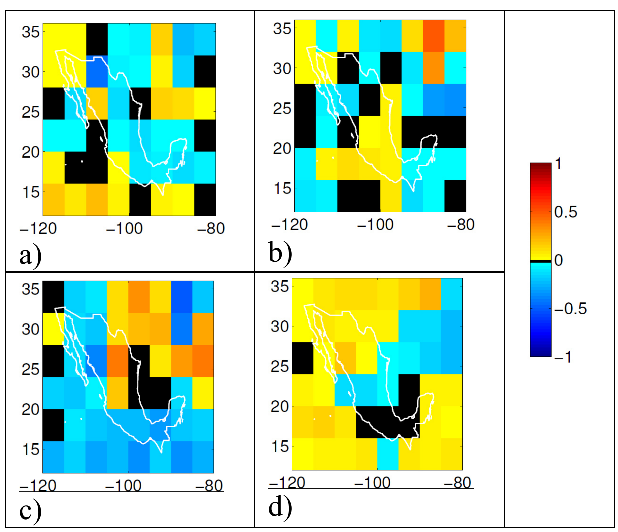

Figure 14.1. Spatial patterns for July precipitation produced by the MZN downscaling method. The maps show the slope coefficient β_1 in Eq. (2) using the GISS-EH model Run 1 under the 20c3m scenario as the predictand, and using as predictors the first 5 principal components (PCs) obtained from (a) Run 2, (b) Run 3, (c) Run 4, and (d) Run 5

Furthermore, do the signs and magnitudes of these estimates have any physical meaning, and do the estimated coefficients express significant relationships? With the exception of a few grid cells in figures 14.2b and 14.2c, none of the estimated coefficients is statistically different from zero. That is, the true value of the parameters is zero even though the point estimates in the map can *randomly* show negative or positive signs, as well as different magnitudes.

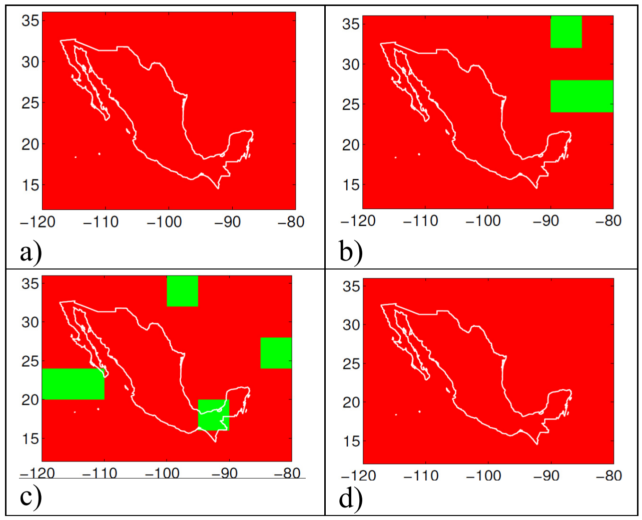

Figure 14.2. Statistical significance of the slope coefficients in Fig. 1. Green areas denote statistical significance at approximately the 5% level ($t > 1.96$ in absolute value).

The results are very similar for the other slope coefficients in Eq. (14.2). The average number of significant slope coefficient values in Eq. (14.2) is close to 5% (Table 14.4), which is the number of false rejections of the null hypothesis expected to occur by chance. Importantly, none of the coordinate points i, j, has a slope coefficient that is significant in all 4 maps (i.e. in different model runs). In addition, the average percentage of significant F values from the estimated regressions is only 3.13%.

The MZN method entails multiplying each grid cell by a random variable with a zero mean and therefore the signs, spatial patterns and magnitudes depend on the sample. Even under the extreme assumption of having a perfect climate model, the downscaling method in MZN can only produce random spatial patterns and magnitudes.

Table 14.4. Percentage of statistically significant coefficients in regression (14.2) at p = 0.05. The predictand is the precipitation field from GISS-EH 20c3m Run 1 and the predictors are the first 5 principal components (PCs) obtained from GISS-EH 20c3m Runs 2 to 5

		β_1	β_2	β_3	β_4	β_5
Predictors obtained from:	Run 2	0	2.1	6.25%	4.2	0
	Run 3	6.3	10.4	0	8.3	6.3
	Run 4	10.4	2.1	2.1	0	16.7
	Run 5	0	12.5	10.4	10.4	4.2
	Average	4.2	6.8	4.7	5.7	6.8

14.2.3.2 What do the regional scenarios in MZN represent?

We extended the results from Section 14.2.3.1 to the case of the CRU database used in MZN. We considered the following regression model:

$$P_{i,j,t}^{CRU} = \alpha + \beta_1 P_{r,t}^{pc1} + \beta_2 P_{r,t}^{pc2} + \beta_3 P_{r,t}^{pc3} + \beta_4 P_{r,t}^{pc4} + \beta_5 P_{r,t}^{pc5} + \varepsilon_t \qquad (14.3)$$

where $P_{i,j,t}^{CRU}$ is the observed July precipitation field from the CRU database for the coordinates i, j; $P_{r,t}^{pc1}, P_{r,t}^{pc2}, P_{r,t}^{pc3}, P_{r,t}^{pc4}, P_{r,t}^{pc5}$ represent, respectively, the first 5 PCs of the simulated precipitation fields obtained from the $r = 1,...,5$ GISS-EH 20c3m runs; as before, ε_t is a sequence of independent and identically distributed random errors such that $\varepsilon_t \sim N(0, \sigma^2)$ and α, β_1 to β_5 are unknown parameters to be estimated.

The upper panels of figures 14.3 to 14.7 show the slope coefficients β_1 to β_5 from regression (14.3), for the 5 GISS-EH 20c3m runs, and the bottom panels show the statistical significance. If the MZN method were correct, the patterns resulting from different runs should be similar. However, this is not the case, as the estimates are realizations of zero mean random variables, and have no physically meaningful relation between large scale and local scale climate variables. The average number of significant cells across model runs for each of the slope parameters is around 5%, only 2.8% are significant in 2 or more runs and none is significant in all 5 model runs.

Finally, the use of an ensemble of downscaled scenarios does not make the climate change signal stronger, nor does it reduce the uncertainty. In the MZN method, the ensemble members are first multiplied by random parameters with zero mean — breaking all original physical patterns and randomly modifying the magnitudes and signs of change— and then a central tendency measure is computed. Evidently, this cannot extract the climate change signal, because probably none of it is left after this downscaling procedure.

14.2.4 Illusory justification and validation of MZN

The abstract of the MZN article begins with the statement 'the climate predictability tool (CPT) is often used to statistically downscale the Intergovernmental Panel on Climate Change scenarios presented in the Fourth Assessment Report'. Taken at face value, this statement —which is not supported by any reference to any published work— would lead the reader to think that the MZN method is in line with the climate change downscaling literature or with the IPCC. This is simply false, and provides no justification for the MZN method.

MZN's own assessment of the performance of their downscaling methodology is limited to subjective comparisons, e.g. stating that the climatology (1971–1999) of the observed fields and that of the downscaled version 'look similar' (MZN, p. 175; their figures 2 and 3). It is trivial to show that these comparisons are irrelevant for assessing model performance: When the explanatory variables are expressed as anomalies and are not dominated by large trends (as is the case), the baseline climatology is represented by the intercept of the transfer function; no virtue can be credited to the model specification other than having included a constant. Finally, the evaluation of trends (MZN, their figures 4 and 5) is not rigorous and based on the authors' subjective perceptions of similarity between maps, without any quantitative or statistical support. Moreover, these figures show some puzzling results, namely that for some regions where both the observations and the model simulation showed the same tendency, the downscaled scenario showed the opposite, suggesting a statistical artifact. Of course, in the case of fundamentally flawed methods such as that of MZN, performance evaluation is pointless.

14.3 Conclusions

The main implication of Estrada et al. (2012a) and the present study is that the regional scenarios produced by the MCZ/MZN downscaling method is erroneous from its roots and its application would invalidate any assessment that uses them as input: from the water availability analyses in MZN to those in the National Climate Change documents of Mexico referenced above as well as others that have been produced since. Our results demonstrate that the publication of MZN and the decision of the Mexican government to maintain online availability of the Magaña (2010) scenarios (http://zimbra.ine.gob.mx/escenarios/) are regrettable and should be of concern not only to Mexico, but to the wider climate change community as well. We stress the urgency of highlighting and remedying MZN's errors before additional misuse of these methods or their products.

Figure 14.3. Estimated values and statistical significance of coefficient β_1 in regression (14.3), using different realizations of the GISS_EH 20c3m experiment. Green areas denote statistical significance at approximately the 5% level ($t > 1.96$ in absolute value). Percentage of cases for which a given cell is significant in more than one map: 2.54%. The percentage of significant cells for each map is given in parentheses.

Run 1 (7.8%)	Run 2 (4.1%)	Run 3 (3.2%)	Run 4 (5.0%)	Run 5 (4.2%)

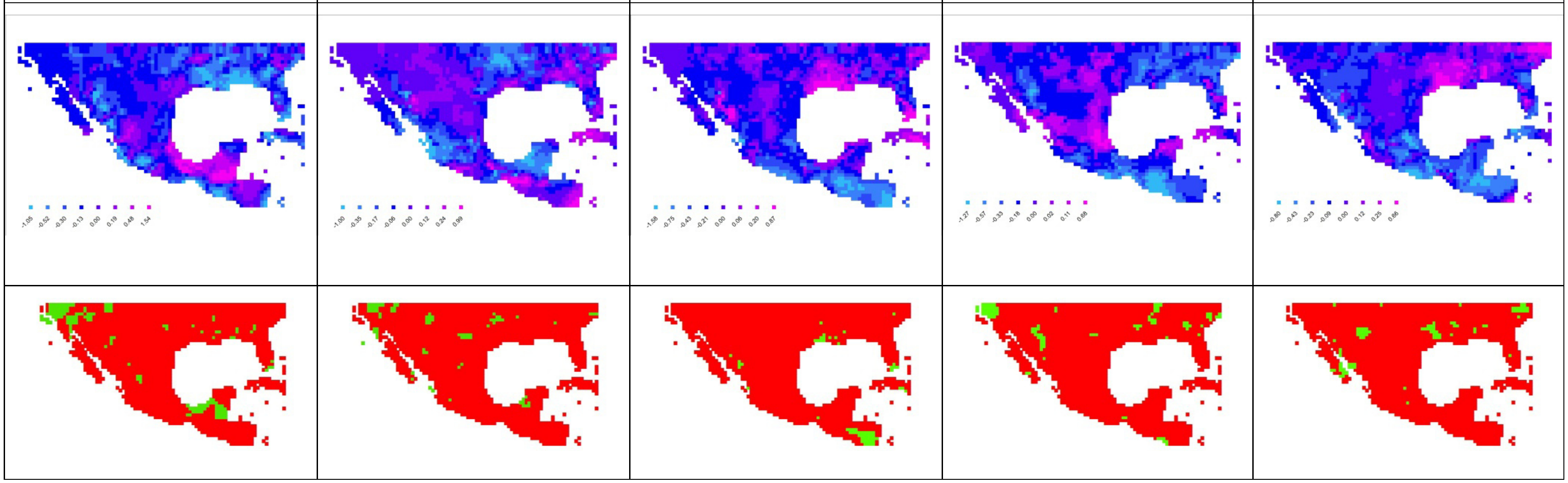

Figure 14.4. Same as Figure 14.3, for coefficient β_2 in regression (14.3)

Run 1 (5.4%)	Run 2 (3.3%)	Run 3 (4.1%)	Run 4 (9.6%)	Run 5 (6.0%)

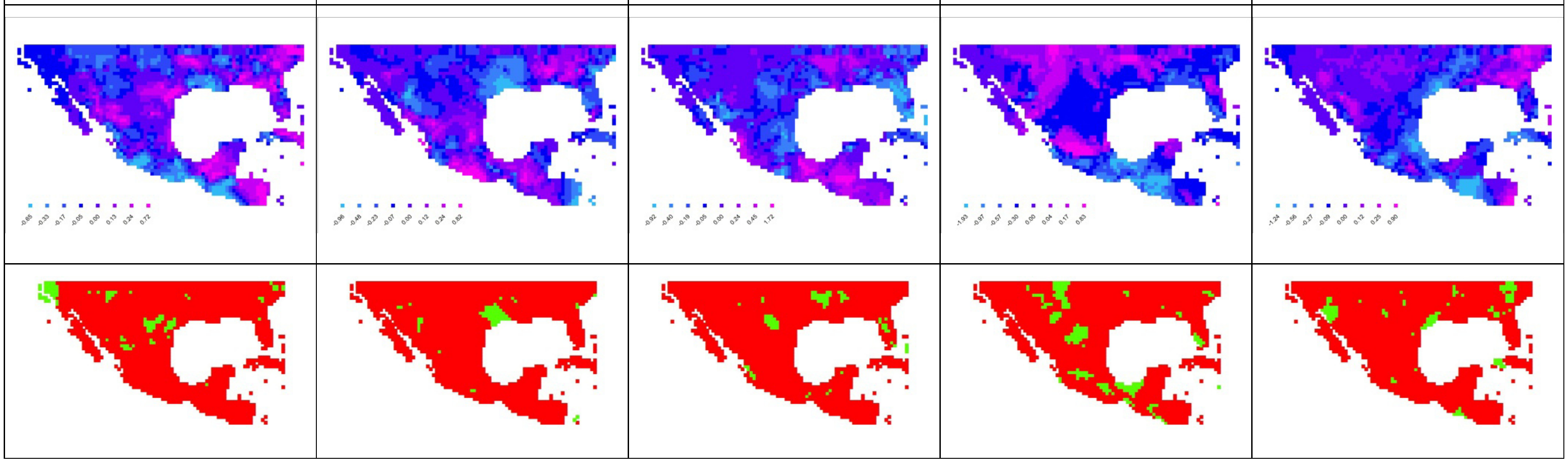

Figure 14.5. Same as Figure 14.3, for coefficient β_3 in regression (14.3)

Run 1 (8.5%)	Run 2 (9.7%)	Run 3 (3.2%)	Run 4 (12.5%)	Run 5 (2.5%)

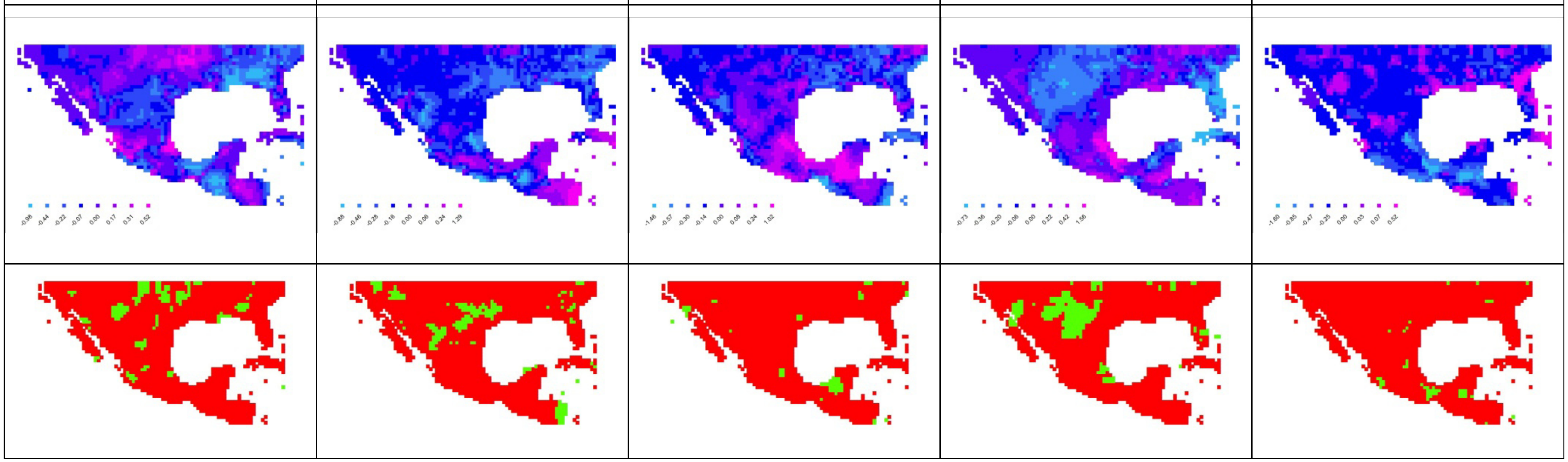

Figure 14.6. Same as Figure 14.3, for coefficient β_4 in regression (14.3)

Run 1 (7.1%)	Run 2 (6.5%)	Run 3 (9.5%)	Run 4 (2.9%)	Run 5 (2.4%)

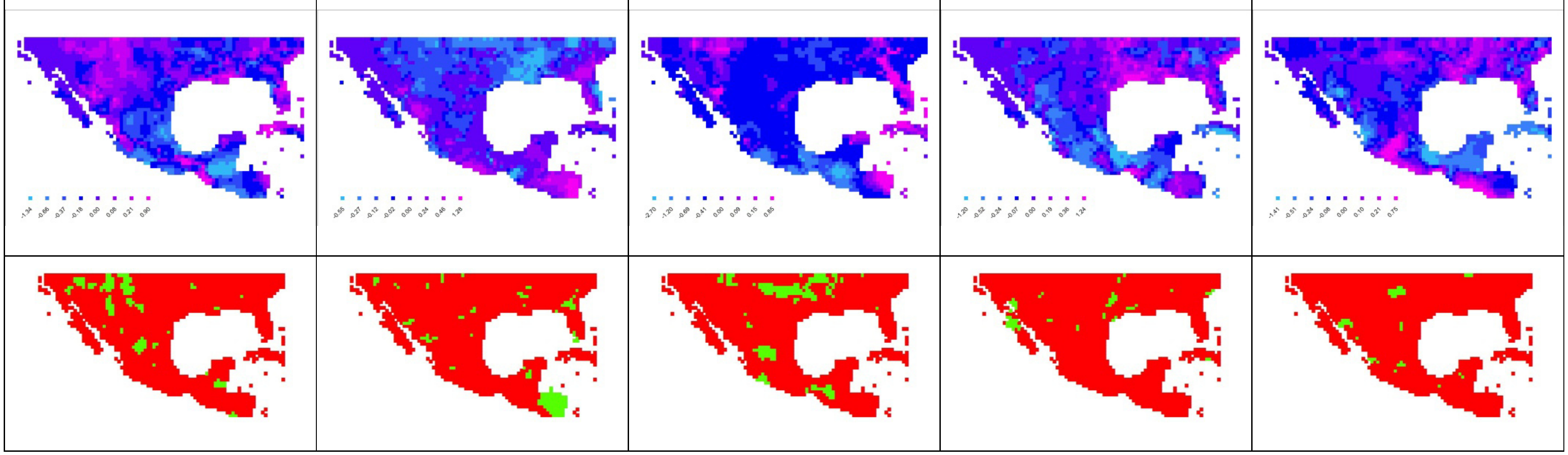

Figure 14.7. Same as Figure 14.3, for coefficient β_5 in regression (14.3)

Run 1 (1.8%)	Run 2 (7.4%)	Run 3 (5.4%)	Run 4 (2.2%)	Run 5 (4.6%)